DOCUMENT RESUME

ED 337 498 TM 017 349

AUTHOR Facione, Peter A.
TITLE Using the California Critical Thinking Skills Test in Research, Evaluation, and Assessment.
PUB DATE 91
NOTE 22p.
AVAILABLE FROM California Academic Press, 217 La Cruz Avenue, Millbrae, CA 94030 ($7.50).
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Tests; College Students; *Critical Thinking; *Educational Assessment; Educational Research; Higher Education; Personnel Selection; Predictive Measurement; Pretests Posttests; Psychological Testing; *Research Needs; Standardized Tests; Student Evaluation; *Test Use; Test Validity; *Thinking Skills
IDENTIFIERS *California Critical Thinking Skills Test (College)

ABSTRACT

The California Critical Thinking Skills Test: College Level (CCTST) is a standardized test that targets core college-level critical thinking skills. It has been characterized as the best commercially available critical thinking skills assessment instrument. Building from CCTST validation studies in 1989 and 1990, this paper proposes avenues for further study and suggests ways that the CCTST might be used. After briefly summarizing the conceptual basis of the CCTST, the paper examines questions from the validation studies, which suggest needed inquiry into the differential impact of typical college-level critical thinking (CT) instruction. Preliminary findings indicate differences among students by academic major and by degree of CT self-esteem. Other findings suggest the need for research into factors that predict student CT ability and characteristics of effective instructors. The use of the CCTST in pretest-posttest studies is considered, given that there is only one form of the CCTST. Strategies for development of local CCTST posttest norms and placement scores are recommended. Possible uses of the CCTST in personnel screening and psychological research are outlined. Five tables present data from previous related studies. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

PETER A. FACIONE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED337498

Using the California Critical Thinking Skills Test in Research, Evaluation, and Assessment

Peter A. Facione
Dean of The College of Arts and Sciences
Santa Clara University
Santa Clara, CA 95053

c. 1991
CALIFORNIA ACADEMIC PRESS
217 LaCruz Ave., Millbrae , CA 94030

TM017349

BEST COPY AVAILABLE

Using the California Critical Thinking Skills Test in Research, Evaluation, and Assessment

Peter A. Facione
Santa Clara University

Introduction

The *California Critical Thinking Skills Test: College Level* (CCTST) is a standardized test which targets core college level critical thinking skills. Praised for the way its multiple-choice items access higher order thinking skills[1] in contexts requiring developed CT dispositions[2], the CCTST has been characterized as the best commercially available CT skills assessment instrument.[3] Building from the CCTST validation studies in 1989/90, this paper proposes promising avenues for further scholarly inquiry and suggests ways the CCTST might be used in research, evaluation, assessment, and placement. After briefly summarizing the conceptual basis of the CCTST, the paper moves directly to questions emerging the 1989/90 findings on CCTST construct validity and concurrent validity. That research suggests needed inquiry into the differential impact of typical college level CT instruction. For instance, we must learn why typical college CT courses appear to advantage certain groups of students over others, as for example men over women. Preliminary findings indicating differences among students by academic majors, and by degree of CT self-esteem also raise challenging research questions regarding typical CT instructional methods. Other findings suggest research into factors predictive of student CT ability and characteristics of instructors which mark them as potentially more effective. Given that there is only one form of the CCTST, this paper addresses the use of the CCTST in pretest/posttest research designs. Strategies for the development of local CCTST posttest norms and placement scores are recommended. The paper outlines possible uses of the CCTST in personnel screening and psychological research.

(c) 1991, Peter A. Facione and The California Academic Press, 217 La Cruz, Millbrae, CA 95093. (Permission to duplicate is available.)

The Conceptual Basis of the CCTST

The *California Critical Thinking Skills Test: College Level* (CCTST) is a 45 minute standardized test designed primarily to assess the core critical thinking skills of post-secondary level persons who are native speakers of English. It was published in 1990 by the California Academic Press in 1990 after more than two decades of conceptual and experimental research. The CCTST is composed of 34 multiple-choice items which target core college level CT cognitive skills. These skills were identified by a national panel of experts who participated for two years in a Delphi research project aimed at achieving an expert consensus regarding what to expect of college freshman and sophomores in terms of critical thinking. The work of this multi-disciplinary national Delphi panel was published in *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction,*[4]

The panel expressed its consensus this way:

> "We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based. CT is essential as a tool of inquiry. As such, CT is a liberating force in education and a powerful resource in one's personal and civic life. While not synonymous with good thinking, CT is a pervasive and self-rectifying human phenomenon. The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. Thus, educating good critical thinkers means working toward this ideal. It combines developing CT skills with nurturing those dispositions which consistently yield useful insights and which are the basis of a rational and democratic society."

The CCTST reports six scores: an overall score on CT cognitive skills and five sub-scores: analysis, evaluation, inference, deductive reasoning and inductive reasoning. Separate percentile norms for college students who have and who have not completed a college level CT course are given. There is ample basis to believe that the CCTST targets a sufficiently rich and contemporary conceptualization of CT, one which is neither esoteric, nor discipline specific. And, the CCTST clearly focuses on the cognitive skills dimension of CT.[5]

Question: Is a focus on CT abilities enough, or must student assessment also take into account more? The national Delphi project identified these affective dispositions as being associated with the ideal critical thinker: Inquisitiveness with regard to a wide range of issues; concern to become and remain generally well-informed; alertness to opportunities to use CT; trust in the processes of reasoned inquiry; self-confidence in one's own ability to reason; open-mindedness regarding divergent world views; flexibility in considering alternatives and opinions; understanding of the opinions of other people; fair-mindedness in appraising reasoning; honesty in facing one's own biases, prejudices, stereotypes, egocentric or sociocentric tendencies; prudence in suspending, making or altering judgments; willingness to reconsider and revise views where honest reflection suggests that change is warranted; clarity in stating the question or concern; orderliness in working with complexity; diligence in seeking relevant information; reasonableness in selecting and applying criteria; care in focusing attention on the concern at hand; persistence through difficulties are encountered; precision to the degree permitted by the subject and the circumstances of inquiry.

Although a large number of CCTST question contexts and distractor choices (wrong answers) invite persons with weak or underdeveloped CT dispositions to make mistakes, the CCTST does not officially purport to target directly these dispositions. It is not unreasonable to suggest, however, that persons whose affective CT dispositions are underdeveloped will not be able to do as well on the CT skills questions. The research opportunity evident here is to determine the extent to which this suggestion is true. **Question: How is the development of a test subject's CT dispositions correlated with that person's demonstrated ability in CT skills as measured by the CCTST?**

CCTST Items and Quantitative Validation Experiments

To determine if an instrument achieves its goal in targeting a given theoretical construct one must go beyond the philosophical to the empirical. A review of the CCTST in terms of face validity reveals that a variety of question formats are employed. Initial items require straight forward **analysis** of a single sentence. Subjects are asked to select the choice that "means the same as" or "is the best interpretation of." The next group of questions require that the roles played by various sentences in a brief paragraph be identified. Is a given sentence part of a reason, is it the main claim or conclusion, is it not

logically relevant to the inference presented? The **evaluation** questions offer short passages and invite the subject to determine the proper inferential strength that the reasons presented lend to the truth of the conclusion drawn. In other questions relating to evaluation a passage is given and an inference draw. Here subjects are asked to evaluate the inference as good or bad and also to state the reason why they have made that evaluation. In the **inference** section questions offer initial sets of statements and invite the subject to indicate what these imply or warrant. Some CCTST question formats resemble those one might find in a reading comprehension test or in the LSAT, SAT, or GRE sections on analytic reasoning. The CCTST concludes with more complex question formats. A passage might include an argument and an objection to that argument. Subjects are asked to evaluate the quality of the objection, indicating if it is a good or bad objection, and giving their reasons for their evaluation. In these situations, as with many of the simpler question formats, deductive and inductive modes of reasoning can be combined, wrong choices based on many different kinds of fallacies can lure uncritical thinkers, and underdeveloped CT dispositions can tend subjects toward wrong choices. [6]

Aside from intuitive judgments about face validity, there are two ways to test empirically whether an assessment instrument hits its target. The most common is by quantitative methods, using experimental and control groups, such as those large scale experiments described below. One assumes that the target phenomenon is present and applies the instrument to see if it is sensitive enough to detect the phenomenon of interest. The second approach is a qualitative variation on this. In the second approach a think-aloud data gathering strategy is used to verify that subjects achieve right answers using good CT and wrong answers using poor CT. If subjects consistently use good CT but make wrong selections, or use poor CT and make correct selections, then the CT instrument's construct validity is questionable.

The CCTST was developed and validated at California State University, Fullerton. The quantitative validation study was conducted in the 1989/90 academic year. Four experiments were conducted to determine if the CCTST was able to measure the growth in CT skills achieved by college students completing approved CT courses. These experiments involved 1169 college students, five courses, three departments, 20 instructors, and 45 sections. The first experiment compared the pretest and posttest means for two independent groups of CT students enrolled in 39 sections of four different campus approved CT courses. The CCTST succeeded in detecting the statistically significant growth in CT skills hypothesized to have resulted from courses approved

specifically for CT instruction. As a control, the second experiment related CCTST score of two independent groups enrolled in six sections of introduction to philosophy. The null hypothesis was retained. In the third experiment, using paired pretest/posttest scores, the CCTST measured the growth in CT skills assumed to have occurred as a result of one semester of approved CT instruction. The fourth experiment retained the null hypothesis for the control group using paired pretest/posttest CCTST scores. Generalizing the results, with a confidence interval of 95%, the range of the mean improvement in the CCTST scores of college students completing approved lower division general education CT courses at public comprehensive universities will be bounded by +1.9071 and +.9861.[7]

How Much CT is Learned in One College Course?

The CCTST reliability coefficient (Kuder-Richardson 20) was .69 on the pretest and .68 on the posttest. These coefficients fall within the .65 to .70 range recommended for tests which purport to target a wide range of CT skills. Of course, one would expect high levels of reliability from a multiple-choice test as compared to an essay test. The theoretical risks of each of these modes of testing are well known, but the actual severity of the pitfalls associated with each is frequently underestimated.[8] The mean number of correct answers out of 34 items on the pretest was roughly 16. A statistically significant increase occurred on the posttest when the mean number correct was just less than 17.

An improvement of hardly one item! Why was the evident growth so small? Initially it would appear that student motivation might have played an important role in diminishing the amount of change. These 1989/90 experiments, including the selection of the Introduction to Philosophy course as the control, were intentionally designed so that everything would be working against the CCTST. Students in the Spring semester '90 pretest samples appeared highly motivated. They were motivated by the good intentions which normally accompany the start of a new semester, they were eager to show that they deserved to be permitted to petition into closed courses, and they gave evidence during the testing sessions of more sustained effort by taking more time to complete the CCTST pretest. The posttest sessions were held during the last week of classes Fall semester and Spring Semester, when students were eager for vacations and yet under the pressure of term paper deadlines and final exams. Students were told their scores on the CCTST did not count for part of their course grades. Posttest students generally seemed to take less time completing the CCTST. Yet, statistically significant growth was detected.

One further indication that the range of improvement recorded in the initial experiments might be smaller than the actual growth in CT is that when the students from two sections of the principle investigator's courses were given the CCTST and motivated by the knowledge that the CCTST was their course final exam, they showed a mean improvement of nearly 5 points from their pretest scores. Naturally the scores of the principle investigator's students were excluded from calculation of the posttest norm.

In contrast to the above concerns about motivation, students who completed the CCTST report finding questions both challenging and interesting. That they were interesting suggests that students found the content sufficiently rich to maintain motivation through to the end of the 34 item test. **Question: What is the actual improvement of students in CT skills which occurs during and as a result of a typical college level CT course?** A design which might resolve this question would be one that used the CCTST in high motivation posttest situations. By examining a sufficiently large number of cases, and controlling for factors such as experimental effects, instructor effects, or discipline effects, a different, and possibly a higher, CCTST posttest norm might be established. See below.

The second method of checking empirically that subjects who take the CCTST arrive at correct choices by way of good CT and wrong choices by way of poor CT uses more qualitative techniques. An important research opportunity exists here. To set the stage for this research one would first require consensus among good critical thinkers on the paradigm patterns of good CT that should (normative/predictive) lead subjects of a given educational level and level of cognitive development to the right answers and the patterns of poor CT that would most likely lead to the selection of wrong answers.[9] This suggests two additional concerns. **Question: How do stages of cognitive development relate to critical thinking skills?** And, **Question: Do subjects at significantly different educational stages exhibit modally different patterns of reasoning?**

Many middle or later CCTST items lend themselves to paradigm analysis. Wrong answers frequently represent faulty reasoning, inattention to data supplied in the question stem, hasty generalization, fallacious thinking, or other mistakes which should be evident to those experts whose CT skills are more refined. Many of the earlier CCTST items which involve more immediate inference or the identification of correct meanings are less amenable to this kind of paradigm analysis. The experts are apt to say of these that a given choice is simply right and another clearly wrong. It is not uncommon for experts

who have internalized so much of their thinking and reading comprehension processes not to be able to articulate how it is that they arrived at the conclusions they find self-evident and beyond the need of further justification. (This is why the sustained contributions of the Delphi experts, as they endeavored to achieve consensus over a period of two years, are so valuable.)

The novice vs. expert distinction may lead to serious problems in the evaluation of the CCTST as well as in the design of this kind of validation research. Since retrospective, reconstructive analyses of one's thinking is notoriously unreliable, large amounts of time will be needed to gather speak aloud data from subjects as they take the CCTST. Fatigue, inter-rater reliability, and sample size, and the order and quantity of think-a-loud questions, all must be considered in designing this kind of research. Replication of this research with persons of different educational levels would be highly useful to confirm or disconfirm theories about purported differences in how persons at various stages of cognitive development think.

The questions raised above, however, do not diminish the one key proposition which the Delphi research and the CCTST validations have firmly and empirically established: **In view of the national Delphi research findings and the quantitative validation studies conducted on the CCTST, we can now assert with very high levels of confidence that those core CT skills which we expect to be part of a college level general education can be taught, learned, and objectively assessed.**

Can We Predict CT Ability or Good CT Instructors?

The 1989/90 CSU Fullerton study included data on over 50 student-related and instructor-related variables. Posttest scores where statistically analyzed using backward multiple regression methods. The three variables remaining in the regression equation when the analysis reached its limits were: SAT verbal, SAT math, and GPA scores, predicting 41% of the variance in the posttest scores. The variables that failed to remain in the equation were the college student's age, units of college work completed, and high school subject matter preparation.[10] **Question: Given that the high school preparation might have fallen out of the regression analysis due to multicollinearity with SAT scores, what exactly is the contribution of factors like this which have strong intuitive validity?**

CCTST results positively correlated with Nelson-Denny reading scores for vocabulary, comprehension, and total score. Non-native English speakers who complete a college level CT course show virtually no change from pretest to posttest. Of six instructor-factors which are hypothesized to be related to effectiveness in teaching CT skills, only years of teaching experience and recent experience teaching CT are related, and these in non-linear ways. Applying the CCTST to the hypothesis that CT skill development is a natural outcome of baccalaureate education, no evidence for that hypothesis, either in general, or by reference to the control groups, could be discovered.[10]

The correlations in Table 1 are based on data from students in the pretest sample, before they have taken any college level CT instruction.

Table 1

Pretest Correlations -- Concurrent Validity

Measure	rho	Sig.	Cases	Mean	Pretest
SAT-Verb	+.55	*p<.000	333	419	16.40
SAT-Math	+.44	*p<.000	333	477	16.40
Col. GPA	+.20	*p<.000	473	2.66	16.11
Reading	.49	*p<.001	42	131.17	17.47
Col. Units	+.03	p=.262	473	66.8	16.11
Age	-.006	p=.449	479	22.03	16.10 [11]

Question: How does the CCTST correlate with other common measures of academic ability or aptitude not used in the 1989/90 research? Research opportunities checking the correlation between the CCTST and other commercially available critical thinking assessment tools, such as the Watson-Glaser, the Cornell, and the Ennis-Weir, would be of interest, provided that the crucial differences in the theoretical constructs each targets are noted.[12] In preparation for post-baccalaureate study professional and graduate schools frequently require applicants to take the LSAT, GRE, MCAT, or GMAT. Correlations with these would be of great interest, particularly since the evidence indicates that baccalaureate level CT instruction increases one's CCTST score. If strong positive correlations exist, then college level CT instruction could be predicted to improve scores on the LSAT, MCAT, etc. Further analysis of the possible correlations with high school GPA in college preparatory courses, and correlations with the PSAT and or state required academic skills tests would be of interest for the findings should have much to tell us about the intended learning outcomes and pedagogy generally used at the high school level. In any such studies appropriate scientific controls for selection, maturation, mortality, experimental effect, Hawthorne effect, and other threats to internal and

external scientific validity must be taken into account. Data on SES, gender, ethnicity, age, academic motivation, and any other variables that might plausibly influence test results should be factored into the analysis.[13]

Question: What factors relating to the CT instructor are associated with greater or lesser student acquisition of CT skills? Research on factors relating to teaching effectiveness has direct implications for hiring practices, faculty evaluation, and staff development. In the 1989/90 CSU Fullerton research, some surprising preliminary findings emerged. No statistically significant relationships to posttest scores were found in the cases of tenure vs. non-tenure status, full-time vs. part time employment status, doctorate vs. non-doctorate preparation, or professor gender. Complex non-linear relationships were suggested in the cases of the number of years of college level teaching experience and the number of CT sections taught in the previous 36 months. Of greater importance is what these findings suggest about those factors which might, indeed, make a difference -- specifically the utilization of particular CT classroom activities, projects, instructional materials, and pedagogy.

Question how can we confirm or disconfirm the emerging intuitive consensus among advocates of CT in the college curriculum that who teaches the CT course is less important than how the CT course is taught? To attack this question controlled experiments should be developed using different teaching techniques in both CT courses and non-CT courses. It may turn out that certain pedagogies, such as active questioning, collaborative assignments, teammate examinations, small group problem solving, and peer-tutoring, work to enhance CT skills even in non-CT courses, whereas other methods, such as lecture, memorization, and homework assignments based on rote drill and practice, do less to enhance CT skill development.

Trying to Determine if We All Teach CT

The issue of effective CT pedagogy is of recurring concern. **Question: What does all this mean for the widely held opinions that we all teach CT and that CT is the natural outcome of a college education?** Contemporary CT advocates, joining richly diverse educational philosophies including American Pragmatism and the Jesuit educational tradition, persuasively argue that CT is a central feature of a solid liberal education.[14] Few would challenge its utility to the individual and to society. In fact, so powerful is the

commitment to teach students to learn to reason well, that many in the profession, regardless of their views about the CT movement, sincerely maintain that this is, indeed, one of the goals they work toward in every course they teach. It is an easy jump from there to the belief that growth in CT is a natural result of a good college education. To evaluate the hypothesis that the baccalaureate experience in general leads to a growth in CT skills it was predicted that the CT of veteran college students would be stronger than those of younger or less experienced students. Operationally, if this were so, then one might predict a positive linear correlation between CT skills and age, or between CT skills and the number of college units earned. However, as indicated in Table 1 above, efforts to discover such results using the CCTST failed.

A second way of trying to test the intuition that all good instruction includes CT instruction was to isolate a specific course, not unlike the required CT course, and determine if a measurable growth in CT skills occurred in that course. For this purpose, and also to make the validation of the CCTST more challenging, Introduction to Philosophy was selected. Intro. to Philosophy, like the four approved CT courses in the 1989/90 CSU Fullerton study, is a lower division general education offering with a student clientele comparable to the cadre enrolled in the approved CT courses. Instructors of Introduction to Philosophy claim with a measure of pride that while teaching CT is not their main goal, they do spend some time, perhaps a week or two, on common fallacies of reasoning. And, more importantly, they emphasize and attempt to model clear and logical thinking throughout the semester.

In the CSU Fullerton study 126 Introduction to Philosophy students took the CCTST under the same controlled conditions as obtained in the Fall semester '89 posttest of the four CT courses. In the Spring semester '90, 124 more students from three matched sections of Intro. Phil. were pretested using the CCTST. The pretest mean was 15.436 and the Fall semester '89 posttest mean was 15.476 revealing a gain of +.04. The t-statistic for this experiment was .08 and the null hypothesis, that there was no significant difference between the two groups, was retained with P=.938. In the Spring semester posttest in May '90 these same three sections were given the CCTST as a posttest. The Spring posttest mean was 16.356 as compared to the pretest mean of 15.722 for the 90 students who complete both the Spring pretest and the Spring '90 posttest. The difference (+.63) is not statistically significant (t-statistic = 1.69, two-tail p=.94). To confirm that the spring and fall groups were reasonably comparable, one could compare the overall Spring '90 posttest mean of 15.722 with the Fall '89 posttest mean of 15.476. The non-significant difference of

0.246 warrants the assumption that the CT skills of these groups are reasonably consistent semester to semester.[7] No "natural growth in CT skills" was evident. **Question: What would a replication of these studies with other control groups of college students find? Question: Would a replication over the length of a year, or two years of general education courses find the same thing?**

CT enthusiasts can justly feel proud that their instructional efforts lead to measurable improvements in students' CT skills. However, it is widely argued in the academy that all good instruction -- almost, but not quite by definition -- does (or should) nurture students' CT skills. Clearly some find it to be an implied criticism to suggest that because CT courses emphasize CT outcomes, other courses in the curriculum do not. In view of the findings presented here, this reaction, however, is inappropriate. Pride in one's teaching does not require that one teach all things. An honest evaluation of one's value as an instructor, or of the value of a course or program of study, should not presume that CT skill development must be an intended outcome. Whether CT should be part of a given course of study is a curricular policy question. Success of achieving CT skill development as an educational outcome is now an empirically testable matter. **Question: How can the CCTST and qualitative CT assessment strategies be used to give evidence of CT growth as a learning outcome of a given course, of a given major or special program, or of the campus general education program?**

Gender, Ethnicity, Academic Major, Language and CT Self-Esteem

Political as well as scientific concerns abound regarding those student-related factors which might enhance or inhibit the development of CT skills at the college level. The validation studies with the CCTST in 1989/90 confirm with confidence that the CCTST does not differentiate unfairly among women and men, nor among people based on their ethnic or racial heritage, nor among students based on their academic majors or level of CT self-confidence. The data with regard to these factors do, however, raise a number of urgent and interesting questions for future research and for CT instruction at the college level and for baccalaureate education in general.

Analyses of pretest data and control group data show that the CCTST is not gender-biased. However, statistically significant gender differences emerge after students completed their college level CT course! Why? Consider Table 2:

Table 2

Differences by Gender

	Men	Women	Difference	Prob.	n-Men	n-Women
High School English	7.65	7.79	-.68	p=.094	272	311
High School Math.	6.53	6.29	.52	p=.091	273	312
SAT-verbal	428	408	-108	*p=.009	288	320
SAT-math	514	459	-18	*p=.001	288	320
College GPA	2.64	2.75	.11	*p=.004	414	263
Pretest	16.3	15.9	-.4	p=.366	237	242
Combined Posttests	17.5	16.7	-.8	*p=.016	328	382
Control Group	15.9	15.2	-.7	p=.214	115	97

At the time of the pretest and among the control group there was no statistically significant difference between the CT skills of men and women. But gender differences were evident when the Fall and Spring semester posttest data were combined. There are two ways the emergence of these differences might be accounted for. The first way is to suggest that the gender differences apparent on the posttest can be attributed to or predicted by the differences in other factors. There is solid evidence to support this. ANCOVA controlling for SAT-verbal and SAT-math scores revealed that gender was not a significant factor in predicting combined posttest variance ($F = .848$; d.f. 1, 464; $p = .358$).[15] This way of accounting for the posttest gender difference suggests that there is something about the scholastic aptitudes that women and men bring to the CT instructional setting which differentially advantage men over women in that setting. On the other hand, perhaps college grading practices and the SAT instrument are gender-biased and men and women do not really bring significantly different aptitudes to the instructional setting. In that case, how can the evident posttest differences be explained?

That a significant gender difference is evident in the combined posttest data suggests that women and men are not acquiring CT skills with equal success in their college level CT courses. **Question: Do men and women have differing expectations for their success in a CT course. Question: Are there differential impacts by gender of the kinds of curricular materials or pedagogical methods typically used in CT courses?** One might, for example, design a study in which one group is taught CT using confrontational and individually competitive instructional settings, where there are winners and losers in classroom arguments. Meanwhile another group might be taught CT using small group, collaborative, and peer-tutoring methods, where cooperation to solve problems is the classroom norm. Comparing the relative growth of the two groups -- using a

pretest/posttest ANCOVA design -- might reveal some interesting things! Other questions to investigate are: **Do the ways in which women and men learn CT differ? If so, how well are these differences understood and accounted for by those who teach CT?**

ANCOVA also indicates that the CCTST does not favor or disadvantage any particular ethnic or racial group. However, not all groups appear to benefit equally from having completed a college level CT course. Consider Table 3, which reports scores for native English speaking students:

Table 3

Differences by Ethnicity/Race of Native English Speakers

	Nat. Am.	Asian	Black	Hispanic	White	Foreign	n	Prob.
Prep-Eng.	n/a	7.96	7.22	7.87	7.88	8.00	444	*p=.001
Prep-math	n/a	6.59	4.90	6.37	6.31	7.20	445	p=.071
SAT-verb	n/a	409	345	421	443	456	474	*p=.003
SAT-math	n/a	480	353	454	498	536	474	*p<.000
GPA	2.83	2.75	2.35	2.54	2.74	2.52	671	*p=.003
Pretest	n/a	16.8	13.0	15.8	16.8	17.6	389	*p=.013
Posttest	15.0	16.7	15.1	16.0	18.1	19.6	502	*p=.002

Table 3 suggests that among native English speakers, blacks (n=13) and foreign students (n=7) registered the largest gains, two points, from pretest to posttest. On average whites (n=395) gained 1.3. The experience of completing an approved college level CT course was not as positive for native English speaking Asians and Hispanics. However one must take into account the statistically significant differences on three factors identified in the regression model described above as predictors of CCTST results. There is a 111 point range in SAT-verbal scores, a 186 point range in SAT-math scores, and range of .48 on college GPA. This strongly suggests that controlling for native language alone is not sufficient to isolate the possible impact of ethnicity/race on CCTST pretest scores. However, **ANCOVA controlling for SAT scores, GPA and native language indicates that ethnicity/race is not a significant factor.** ANCOVA were run on CCTST pretest scores, Fall semester posttest scores and combined Fall and Spring posttest scores. In no case was ethnicity a significant factor when SAT scores, GPA, and native English language ability were controlled factors.[15]

How do students from different college disciplines do on the CCTST? Presented with the prompt "The major in which I hope to graduate can best be grouped with..." students were given six clusterings of majors from which to select one. The six were formed on the basis of the epistemological and methodological similarities and differences

hypothesized by this researcher to obtain among the disciplines in each cluster. Table 4 indicates the pretest and combined posttest results for each of the six. Fortunately every group appears to benefit from CT instruction. However the benefits do not appear to be equally distributed. While academic major was not a significant factor on the CCTST pretest, scores on the posttest did vary significantly by major. Indeed, ANOVA of the posttest results indicate that academic major (as here clustered) is a statistically significant factor with regard to CCTST performance, ($F = 5.2253$; d.f. 6, 719; $p = .0000$). However, academic major was not statistically significant with regard to the CCTST pretest, ($F = 1.4661$; d.f. 5, 468; $p = .1995$).[15] **Question: Why do students from different majors appear to start or end their CT courses with different growth patterns? Question: As with the gender issue, are their predispositions, learning styles, or pedagogical differences at work?**

Table 4

CCTST Differences by Grouped Academic Majors

Group and % of Cases	Pretest	Posttest	Change
A. Letters, languages, English, Liberal Studies, History, Humanities. [18%]	17.18	18.50	+ 1.32
B. Social Sciences, Psychology, Human Services, Teaching. [20%]	15.82	16.93	+ 1.11
C. Mathematics, Engineering, Statistics, Computer Sci. [9%]	16.14	18.18	+ 2.04
D. Natural Sciences, Physical Sci., Health Professions. [7%]	16.77	16.86	+ .09
E. Business, Administration, Management, Government, Military Science. [39%]	15.80	16.43	+ .63
F. Performance Studies, Drama, Art, Music, Physical Ed. [6%]	15.47	16.19	+ .62
Z. Omit -- No response [<1%]			

Contrasting Outcomes with Student Perceptions

The strong positive correlation of CCTST with college GPA mentioned earlier does not match the students' perceptions. When the pretest group was asked to respond to the statement: "My GPA is an accurate reflection of how logical my thinking is," 224 students (47%) indicated "No, not really," and 170 (35) said "More yes than no." Only 49 (10%) said "Yes it is," whereas 34 (7%), indicated "No, they do not match at all." These misgivings about the relationship between their GPA and their CT ability might be

attributable to uncertainty on the part of pretest students regarding what CT was. One might expect, therefore, that after having completed an approved CT course, their perceptions about the relationship between their GPA and their CT ability might have changed. But they did not. Given the same prompt, on the Fall semester posttest, 42% (196 of 465) said "No, they do not match," 35% (161) answered "More yes than no," 14% (65) said "Yes it is," and 9% (41) responded "No, they do not match at all."[15]

Question: Why do students perceive their GPA and their CT abilities not to be strongly correlated when in fact they are? In designing a study around this issue it might be useful to examine students' views about what CT is and whether it is useful. It might also help to find out if students are generally skeptical about the GPA. Something important about how students view college level learning might well be working here, if we can find out what it is.

To explore their CT self-confidence, students were asked to respond to the prompt, "Critical thinking and being logical are quite easy for me." Of the 480 pretest students 383 (80%) gave positive responses and only 96 gave a negative response. On the Fall semester posttest 392 (84%) gave positive replies and only 72 of 465 were negative. This level of CT self-confidence at posttest time seems particularly surprising, if not entirely unjustified, considering that the 16.83 posttest mean represents only 49.5% correct out of 34 items.[15] **Questions: Given what might be described as the "CT over-confidence" of these students, what is the basis for their self-assessments? What have we educators, or others, done to promote in college students the notion that they should feel good about having a set of cognitive skills which, when exercised, yield the correct results only about half the time?**

Table 5

CT Self-Confidence and CCTST Scores

Response	N	Pretest	N	Fall Posttest	N	Spring Posttest
A. Yes, to be honest it is.	107	17.41	149	18.83	60	19.65
B. Well, I sort of agree.	276	16.36	243	16.63	148	16.80
C. No, not really.	86	14.21	67	14.93	48	16.46
D. Are you kidding.	10	11.40	5	14.40	6	16.67

Establishing Local CCTST Norms and Cut Scores

Technical Report #4 provides pretest and posttest percentile norms for the CCTST and for each of its five sub-tests. These norms are based on the analysis of 1673 test forms

completed by representative samples of college students during the 1989/90 academic year at a comprehensive, urban, state university. 781 cases were used to from the pretest norm and 892 to form the posttest norm. Out of a possible 34, on the pretest the scores ranged from 2 to 29, and on the posttest from 3 to 31. The pretest mean was 15.89 and the posttest is 17.27.[5] The students in these groups averaged 900 on their combined SAT-verbal and SAT-math scores. The mean age was 22 years. The students studied had typically completed enough semester units to qualify for junior standing, even though the CT requirement was a lower division general education requirement. Typically these students had completed nearly four years of high school preparatory English (7.8 semesters), and just a bit more than three years of high school preparatory Math (6.4) semesters. Many institutions find that their populations are close enough to the sample such that they can use the norms provided in Technical Report #4 without modification.

But there might be reasons to modify the norms, particularly if the CCTST is used as a way to exempt students from a CT course. First it should be remembered that nearly 19% of the students in this sample were non-native English speakers, and their mean scores on both the pretest and the post-test were 13.75.[10] The pretest mean for a sub-group of 472 students showed that the 388 who were native English speakers had a mean score of 16.65. Looking at 462 posttest students, the 373 in this group who were native English speakers had a mean score of 17.59. Second, as indicated above, there is reason to believe that student motivation for the posttest was less than optimal, which would suggest that the posttest norms might be low.

Institutions with a CT requirement may use the CCTST as a placement test to exempt certain students from that requirement. In addition to being reasons why an institution might wish to consider creating its own posttest norms, the two considerations above suggest that the CCTST cut score for course exemption purposes might reasonably be set higher than at 17, which was the modal score on the CCTST posttest. Different strategies can be used to establish local norms of the CCTST, particularly if it is used as a tool for purposes of course exemptions. One is to determine which percentage of the population the institution might reasonably wish to exempt. For example, if the policy decision is that 25% of the entering freshman class probably would not have to be required to take a CT course, then the local 75th percentile cut of score could be used. The students are given the CCTST and those in the institution's top CCTST quartile are considered to have satisfied their CT course requirement. Another strategy is to create local posttest norms after carefully testing large numbers of students who have completed

an institution's CT program. Assurances that a wide variety of sections are used to achieve this would be necessary to control for the specific strengths or weaknesses of different instructors or different disciplinary approaches to CT. However, the two strategies do not mix well. An institution would not want to create its own posttest norm using a sample population out of which the top 25% had already been removed.

To establish local norms a sample of at least 500 is recommended. Percentile scores associated with each possible number of correct answers on the CCTST are easily derived. Percentile scores provide an ordinal ranking which can be misleading if the sample upon which there were derived is too small or not normally distributed. For smaller samples or for samples that not normally distributed, it is recommended that percentile scores be converted to transformed normalized standard scores (T-Scores) before parametric statistical analysis and interpretation is undertaken.

Posttest Only and Pretest/Posttest Research Designs

For the present, the CCTST comes in only one form.[16] This provides advantages from the point of view of statistical analysis, particularly in that questions regarding multi-form reliability and equivalence are moot. Where there is reason to suspect a possible testing effect, posttest only research designs are used. In that way both students and instructors can remain blind to factors which knowledge of a pretest might contribute to one's preparation for a posttest. Program evaluation using posttest only design is a legitimate paradigm for inquiry into the effectiveness of a mode of instruction or pilot curriculum in terms of predetermined learning outcomes. The CCTST is particularly well suited to be used in posttest only program evaluation research.

Some educational research strategists recommend the pretest/posttest design for program evaluation and student assessment because it permits the use of ANCOVA analysis of gain scores. The initial impetus to use one's entire sample, giving everyone a pretest and a posttest, is not necessarily the only way to design such research. Groups can be divided in half by matched pairs and one member of each pair can be given a pretest while the other is later given a posttest. Intact sections of courses which result from the random assignment of students can be treated as control and experimental groups and, if multiple sections are available -- as should be the case to control for instructor effectiveness variance -- different sections could be given either a pretest or a posttest.

In those designs requiring that the same student be tested more than once, the CCTST can be used effectively without introducing the problem of inter-form reliability that is invariably created whenever alternative forms of the same instrument are used. The research in 1989/90 showed no testing effect in the elapsed time of one semester. That the control group showed no statistically significant gain, when in fact the same form of the CCTST was used in both the pretest and the posttest of one of the control group samples, is a preliminary indication of this. That there was no statistically significant difference between the posttest of the control group which received the CCTST as a pretest and the control group sample which had no pretest, is another indicator that the CCTST does not have a instrumentation effect that carries over the length of a semester. Anecdotal responses from students consistently carry two messages: The CCTST is interesting to take. And although the topic content of questions can be remembered on retaking the test, the answers first selected cannot be remembered and are not perceived as useful recollections.

Other Adaptations of the CCTST

The CCTST has drawn the interest of personnel officers and persons screening applicants for positions which require a measure of independent problem solving and decision making. On the hypothesis that good critical thinkers would be better suited for administrative and executive positions than poorer thinkers, the CCTST has been used as a preliminary screening instrument. In such cases the pretest norms are used to get a rough indication of where a candidate might be relative to others. That age is not correlated with CCTST results is a further indication that this use of the CCTST is not biased by the age of the person asked to take it. Considerations regarding native language and reading comprehension, however, should not be overlooked. While CCTST results might be used as one possible indicator of potential success in positions requiring stronger critical thinking ability, the CCTST score should not be the only factor in determining the qualifications of candidates.

Psychological research, as for example in nursing, medical anthropology, or economics, into the factors which influence persons to make certain decisions or act in certain ways, when these decisions or actions have strongly cognitive bases, can fruitfully use the CCTST. To determine the possible effect of the subjects critical thinking ability on

the decision or behavioral outcome variable, the CCTST can be administered and its scores used in multiple regression analyses. Research using theoretical models which provided a role for intelligence, metacognition, cognition, problem solving, or decision making, might find the theoretical construct of the CCTST better suited to the inquiry.

Research that targets a subject's reasons for selecting a particular answer choice can be built into the CCTST. For example, when taking the CCTST subjects can be directed to write a one sentence explanation for why they selected a given choice on a given question. This way the CCTST items provide opportunities for assessors to look at the product of a subject's thinking in the commentary and explanation that results from the exercise of the core CT skills tested on the CCTST. If written responses to CCTST items are of interest, it is recommended that only selected questions be used and that reasonable adjustments be made in the time permitted.

Endnotes

[1] Robert Ennis, commending on the CCTST at the Central Division meetings of the American Philosophical Association in Chicago in March 1991, indicated that the question contexts, particularly those in the later part of the CCTST, were rich in opportunities to distract persons with underdeveloped CT dispositions and those susceptible to various forms of fallacious reasoning.

[2] Michael Scriven, at the Eleventh International Conference on Critical Thinking at Sonoma, California, August 1991, used the multiple-choice items on the CCTST to exemplify the best test items of their kind. In his remarks, made in the context of a comparative evaluation of multiple-choice items, multiple-rating items, and essay test items, Scriven judge the items on the CCTST as being more capable of accessing higher order thinking skills that the items on the LSAT.

[3] JoAnn Carter-Wells, at the Eleventh International Conference on Critical Thinking at Sonoma, California characterized the CCTST as the best commercially available CT assessment instrument. Her evaluation came in the context of a comparative analysis of four CT tests (the CCTST, Watson-Glaser, Ennis-Weir, and Cornell) using a specially designed matrix of criteria for test evaluation.

[4] Facione, Peter A., Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction, (ERIC ED 315 423), The California Academic Press, Millbrae, CA, 1990. (80 pages with appendices.) A 22 page

[5] Facione, Peter A., "Technical Report #4, Interpreting the CCTST, Group Norms and Sub-Scores," (ERIC TM 327 566), The California Academic Press, Millbrae CA, 1990. Technical Report #4 provides norms for each of the five CCTST sub-tests used either as a pretest or a posttest. Three sub-tests mirror the Delphi conceptualizations in targeting the following theoretical constructs:

Analysis as used on the CCTST has a dual meaning. First it means "to comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures or criteria," which includes the sub-skills of categorization, decoding significance, and clarifying meaning. Analysis on the CCTST also means "to identify the intended and actual inferential relationships among statements, questions, concepts, descriptions or other forms of representation intended to express beliefs, judgments, experiences, reasons, information or opinions," which includes the sub-skills of examining ideas, detecting arguments, and analyzing arguments into their component elements.

Evaluation as used on the CCTST has a dual meaning. First it means "to assess the credibility of statements or other representations which are accounts or descriptions of a person's perception, experience, situation, judgment, belief or opinion; and to assess the logical strength of the actual or intended inferential relationships among statements, descriptions, questions, or other forms of representations," which includes the sub-skills of assessing claims and assessing arguments. Evaluation on the CCTST also means "to state the results of one's reasoning; to justify that reasoning in terms of the evidential, conceptual, methodological, criteriological and contextual considerations upon which one's results were based; and to present one's reasoning in the form of cogent arguments" which includes the sub-skills of stating results, justifying procedures, and presenting arguments.

Inference as used on the CCTST means "to identify and secure elements needed to draw reasonable conclusions; to form

BEST COPY AVAILABLE

conjectures and hypotheses, to consider relevant information and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, questions, or other forms of representation," which includes the sub-skills of querying evidence, conjecturing alternatives, and drawing conclusions.

The two other sub-tests on the CCTST follow more traditional conceptualizations which divide the realm into inductive and deductive reasoning. These concepts, however, have become notoriously ambiguous as a result of important differences in what they denote in different disciplines. Concern about this ambiguity explains why the words "deduction" and "induction" appear nowhere in the CCTST. However, in view of the continued use of this distinction, the CCTST offers these two sub-scores. Following the lead of contemporary logicians, the CCTST grounds its conceptualization of the deductive vs. inductive distinction on the basis of the purported logical strength of the inference.

Deductive Reasoning as used in the CCTST sub-score means the assumed truth of the premises purportedly necessitates the truth of conclusion. Not only do traditional syllogisms fall within this category, but algebraic, geometric, and set-theoretical proofs in mathematics (including "mathematical induction") also represent paradigm examples of deductive reasoning. Instantiation of universalized propositions is deductive, as are inferences based on such principles as transitivity, reflexivity and identity. In the case of valid deductive arguments, it is not logically possible for the conclusion to be false and all the premises to be true.

Inductive Reasoning as used in the CCTST sub-score means an argument's conclusion is purportedly warranted, but not necessitated, by the assumed truth of its premises. Scientific confirmation and experimental disconfirmation are examples of inductive reasoning. The day to day inferences which lead us to infer that in familiar situations things are most likely to occur or to have been caused as we have come to expect are inductions. Statistical inferences are inductive, even if the inference is the prediction of an extremely probable specific (rain today) based on general principles (meteorological laws) and a given set of observations. Inference used to inform judgment by reference to perceived similarities or applications of examples, precedents, or relevant cases, such as is typical of legal reasoning, is inductive. Also inductive is that common and powerfully persuasive -- even if logically suspicious -- tool of everyday dialogue, analogical reasoning. In the case of a strong inductive argument it is unlikely or improbable that the conclusion would actually be false and all the premises true, but it is logically possible that it might.

[6] Facione, Peter A., "Assessing Inference Skills," (ERIC TM 012 917), 1989

[7] Facione, Peter A., "Technical Report #1, Experimental Validation and Content Validity" (ERIC ED 327 549), The California Academic Press, Millbrae CA, 1990.

[8] Facione, Peter A., "Thirty Ways to Mess Up A CT Test," Journal of Informal Logic, Vol. 12, No. 2., Spring 1990, pp. 106-11.

[9] Dr. JoAnn Carter-Wells, Dept. of Reading, CSU Fullerton, is conducting some very interesting research along these lines using audio taped think-a-loud sessions during which college students work through various CCTST question items. Dr. Wells is also looking for evidence of cognitive shifts in problem solving strategies, and the relationships between critical thinking and critical reading.

[10] Facione, Peter A., "Technical Report #2, Factors Predictive of CT Skills" (ERIC ED 327 550), The California Academic Press, Millbrae CA, 1990.

[11] The Measures are: Scholastic Aptitude Test-Verbal, SAT-Math, Nelson-Denny Reading Test total score, College Grade Point Average on a 4 point scale, Number of semester units of college work earned, and age in years.

[12] At Hartwick College in New York, Ms. Judith Rulund is pretesting and posttesting the 1991/92 freshman class with the CCTST and the Watson-Glaser to determine what relationships might exist.

[13] At the University of Kentucky, Lexington, Mr. Patrick Keenist, Sociology Dept., is examining the relationship between SES and CT abilities of college students across ethnic groups. In related research Mr. Alvin Y. Wang, Psychology Dept., University of Central Florida, is examining cultural-familiar predictors of children's metacognitive and academic performance. Preliminary findings from Mr. Wang's research suggest that ethnicity and race drop out as non-factors as compared to socioeconomic considerations.

[14] Facione, Peter A., "Critical Thinking: What It Is and Why It Counts," The California Academic Press, Millbrae, CA., 1991.

[15] Facione, Peter A., "Technical Report #3, Gender, Ethnicity, Major, CT Self-Esteem, and the CCTST," (ERIC ED 326 584), The California Academic Press, Millbrae CA, 1990.

[16] The conceptually equivalent second form is expected in August 1992.

BEST COPY AVAILABLE